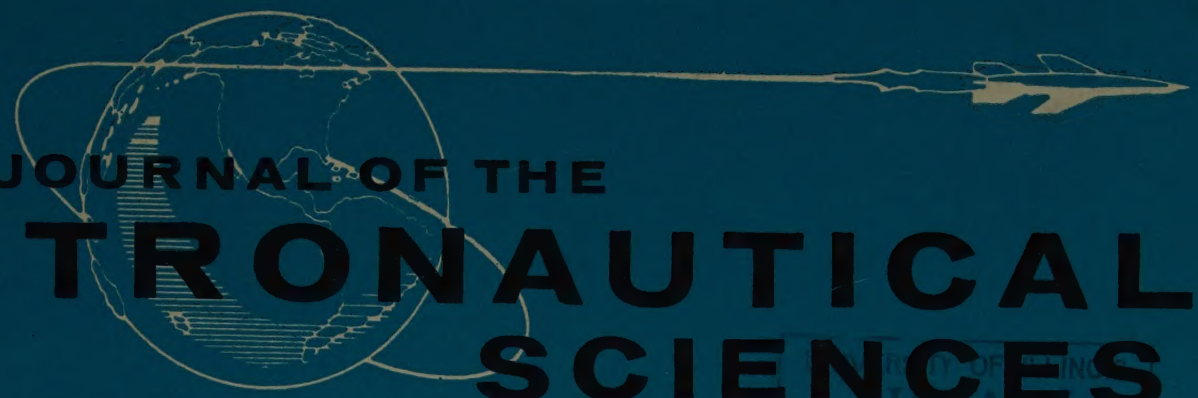


THE JOURNAL OF THE ASTRONAUTICAL SCIENCES

VOLUME VI, NO. 4 WINTER 1959

CONTENTS

PUBLISHED BY THE AMERICAN ASTRONAUTICAL SOCIETY, INC.

THE AMERICAN ASTRONAUTICAL SOCIETY, INC.

516 Fifth Avenue, New York 36, New York, U.S.A.

THE AMERICAN ASTRONAUTICAL SOCIETY

The American Astronautical Society, founded in 1953 and incorporated in New York State in 1954, is a national scientific organization dedicated to advancement of the astronautical sciences. The society considers manned interplanetary space flight a logical progression from today's high-performance research aircraft, guided missile, and earth satellite operations. The scope of the society is illustrated by a partial list of the astronautical fields of interest: astronavigation, biochemistry, celestial mechanics, cosmology, geophysics, space medicine, and upper atmosphere physics, as well as the disciplines of astronautical engineering, including space vehicle design, communications, control, instrumentation, guidance, and propulsion. The aims of the society are to encourage scientific research in all fields related to astronautics and to propagate knowledge of current advances. Promotion of astronautics in this way is accomplished by the society largely through its program of technical meetings and publications.

AFFILIATIONS

AAS cooperates with other national and international scientific and engineering organizations. AAS is an affiliate of the American Association for the Advancement of Science and a member organization of the International Astronautical Federation.

MEMBERSHIP REQUIREMENTS

All persons having a sincere interest in astronautics or engaged in the practice of any branch of science, which contributes to or advances the astronautical sciences, are eligible for one of the various grades of membership in the Society. Requirements are tabulated below. A special category of Student Membership has been authorized for full time students or those under 18 years of age. A nominal membership fee of $5.00 is made in such cases to cover publication costs. The Directors of the Society may elect as Fellows of the Society those who have made direct and significant contributions to the astronautical sciences. Information regarding individual membership as well as Corporate and Benefactor Membership may be obtained by writing the Corresponding Secretary at the Society address.

Grade	Contribution To Astronautics	Experience or Scientific Training*	Annual Dues
Affiliate Member	Interest	none required	$8
Associate Member	Direct Interest	4 years	$10
Member	Active Interest	8 years	$10
Senior Member	Recognized Standing and Direct Contribution	10 years	$15

* A Bachelor's, Master's or Doctor's degree in any branch of science or engineering is equivalent to four, six or eight years of experience, respectively.

JOURNAL OF THE ASTRONAUTICAL SCIENCES
Director of Publications, Ross Fleisig
Editor, Robert E. Roberson
Associate Editor, Charles H. Moss
Assistant Editor, Carl A. DuNah, Jr.
Circulation Manager, George Clark
Address all Journal correspondence to Box 24721, Los Angeles 24, Calif.

Subscription Rates: One year $5.00; foreign $6.00; single copy $1.25. The Journal is published quarterly and sent without charge to members of the Society.

Integration Methods for Differential Equations

S. Schlesinger†

Abstract

This paper describes a series of numerical experiments with integration techniques for differential equations with an aim to develop criteria for choosing numerical integration methods for astronautical trajectory and orbit computations.

The requirements of these astronautical integrations impose extreme restrictions on the integration techniques, from the point of view of accuracy and the time required for the integration. Since accuracy in numerical integration processes is usually obtained only at the expense of speed of integration, and speed of integration obtained usually at the expense of accuracy, this paper discusses integration methods that enable one to make the best of this inevitably difficult situation.

Introduction

This discussion deals with various methods of numerical integration of differential equations with particular concentration on second order differential equations in which the first derivatives are absent (as in the case of drag-free motion in a gravitational field). The techniques described herein are presented in standard numerical analysis texts; however the numerical comparisons included in this paper are not found in such sources.

Methods that were considered can be placed in three general categories:

1) Difference techniques (including sum-difference).
2) Runge-Kutta and its variations.
3) Predictor-corrector techniques.

For the methods involving difference tables, a measure of accuracy can be gained by using methods which are specifically suited to second order equations with first derivatives absent. Such a method is the Second Sum (or Gauss-Jackson) Method, which is a variation of a direct difference scheme involving only differences of the second derivative. This technique has been used with great success by astronomers for many years. A heuristic discussion of the advantages of this technique can be found in reference [1] and its derivation can be found in [2].

The Runge-Kutta method that was studied was the standard fourth order method [3]. The Gill modification of Runge-Kutta was not studied since the saving in storage space that would result is insignificant in a medium or large size computer, and the gain in accuracy from less round-off error is insignificant, since Runge-Kutta should not be used on such a long computation that round-off could become a prominent error source. The advantages of Runge-Kutta methods make the procedures very useful for starting an integration and for integrating up to specified positions in the independent or dependent variables, but the computation time per step clearly prohibits the use of these methods for long computations.

Predictor-corrector techniques were investigated with primary emphasis on a technique which is specifically suited to second order differential equations with first derivatives absent [4].

As will be indicated, this method proved to be the most disappointing of those studied since it suffered both accuracy and stability troubles when compared with the other two categories of integration techniques. Recent modifications of these predictor-corrector methods have been developed to overcome the instability problem [6]. However these methods reduce the accuracy of the integration (which was not particularly good in original form).

Difference Table Techniques

Probably the fastest integration methods rely upon a table of differences (and possibly sums) of derivative values. These techniques do not have the flexibility of convenient interval change that is characteristic of Runge-Kutta methods, but they do involve significantly less computation time to perform a comparable integration provided a uniform integration interval is advisable for the numerical integration.

The most generally accepted integration methods of the difference table type are Adams Method and the Adams-Bashforth Method [5]. The second method is actually a variation of the first one in which a correction formula is used to improve the accuracy (at the expense of doubling the computing time).

The Second Sum Method is compared with both the Adams Method and the Adams-Bashforth Method, and also with variations on itself in which a correction formula is introduced to improve the accuracy of the integration (and double the computation time).

A series of comparisons is made between these methods in the next section, as well as comparisons with other classes of methods. Unless specifically indicated the Second Sum Method will not involve the correction formula which doubles the computing time. In viewing these tables and subsequent conclusions, it should be borne in mind that the normal Second Sum and Adams Method require about the same computing time and the *corrected* Second Sum Method and Adams-Bashforth (which is simply a corrected Adams Method) each take about twice as much computing time as Adams or the uncorrected Second Sum Method.

† Manager, Mathematical Services, Aeronutronic Systems, Inc., Newport Beach, California.

Prior to comparing these difference table techniques with the other methods studied in this report, we will first investigate some comparisons between Adams, Adams-Bashforth, and Second Sum Methods. Following this, we will consider the question of whether or not a corrector should be used with the Second Sum Method. These studies will utilize numerical examples of the solution of three differential equations which, as will be justified in the next section, test for truncation error, instability, and a combination of the two effects.

Solution of $\frac{d^2y}{dx^2} = -y$ *(Testing Truncation Only)*

x	Analytic Solution $\sin x$	Adams 4∇, $\Delta t = .2$	Adams-Bashforth 4∇, $\Delta t = .2$	Σ^2 4∇, $\Delta t = .2$
5	−.95892427	−.95*929783*	−.95892*849*	−.95892*320*
8	.98935825	.990*02624*	.98937*139*	.98935*612*
14	.99060736	.99*173687*	.990*65144*	.99060*368*

In this table, 4∇ represents a numerical solution including all differences up to 4th differences and Σ^2 indicates the Second Sum Method. The underlined figures represent the figures that are in error and give a crude but graphic indication of the error. These abbreviations and conventions will be used throughout this section and the next section.

Notice that the Second Sum Method is superior to both Adams and Adams-Bashforth Methods despite requiring much less time than the Adams-Bashforth Method and about the same time as Adams Method. Unfortunately, this situation is reversed when instability arises.

Solution of $\frac{d^2y}{dx^2} = y$ *(Testing Instability only)*

x	Analytic Solution e^{-x}	Adams 4∇, $\Delta t = .2$	Adams-Bashforth 4∇, $\Delta t = .2$	Σ^2 4∇, $\Delta t = .2$
5	.67379470 (−2)	.67*363983* (−2)	.67379*618* (−2)	.67381*116* (−2)
7	.91188197 (−3)	.91*305590* (−3)	.91188*589* (−3)	.91*311452* (−3)
11	.16701701 (−4)	.*31340425* (−4)	.16701*842* (−4)	.*84087840* (−4)
27	.18795288 (−11)		.18795*768* (−11)	

In this example, it is rather clear that the Adams-Bashforth computation is far superior to either of the other methods, even when the additional computing time is taken into account. The fact that the error does not grow, even for an extensive computation, indicates that the *Adams-Bashforth Method is excellent for integrations in which instability is a problem.*

To study the interaction between instability and truncation error, a combination of exponential and sinuosoidal functions was used in an attempt to evaluate these methods.

These results are rather disconcerting in the sense that none of the methods could successfully integrate the third test differential equation for the specified interval size. What is significant is that the Second Sum Method was no worse than Adams-Bashforth (which was only slightly better than Adams Method). On this basis plus the first test case, it is a rather obvious conclusi that the *Second Sum Method is excellent for integrati in which truncation error due to function oscillation is dominant error source.*

Solution of $\frac{d^4y}{dx^4} = -4y$ *(Testing Instability and Truncation*

x	Analytic Solution $e^{-x} \sin x$	Adams 4∇, $\Delta t = .1$	Adams-Bashforth 4∇, $\Delta t = .1$	Σ^2 4∇, $\Delta t = .1$
1	.30955987	.30955*477*	.30956*019*	.30955*989*
3	.70259519 (−2)	.70*301590* (−2)	.70*261810* (−2)	.70*264538* (−
5	−.64611811 (−2)	−.64*631293* (−2)	−.64*629283* (−2)	−.64*627829* (−
8	.33189276 (−3)	.*36497755* (−3)	.*36141908* (−3)	.*35411446* (−
11.2	−.13389468 (−4)	−.*87215301* (−3)	−.*79004810* (−3)	−.*66102331* (−

Since Adams-Bashforth is simply a modification Adams Method in which a corrector formula is use and since this corrector greatly improves the stabili of the integration method, it would seem that a co rector formula should be used with the Second Su Method. Because the Second Sum Method is weak on in its ability to control instability, this would see to be a method of obtaining an excellent general pu pose integration technique.

In order to retain the basic advantage of the Seco Sum Method in that it does not use the first derivati terms (which are absent from the differential equatio of the form being studied), the corrector formula based on the fundamental central difference formula:

$$\delta^2 y_0 = h^2 \left(y_0'' + \frac{1}{12}\delta^2 y_0'' - \frac{1}{240}\delta^4 y_0'' + \frac{31}{60{,}480}\delta^6 y_0'' - \frac{289}{3{,}628{,}800}\delta^8 y_0'' + \cdots \right.$$

This formula is actually used as the basis of the Seco Sum Method and all of the correctors. The second diff ence corrector, abbreviated δ^2, uses it directly, the fi sum corrector, abbreviated Σ, uses the formula af applying the sum operator once, and the second su corrector, abbreviated Σ^2, uses the formula after appl ing the sum operator twice.

The results obtained from using these correctors summarized in the following tables. The same thr differential equations used previously were the basis comparison. Unfortunately, the correctors are rath disappointing in their effect on these trial solutio sufficiently so that the use of a corrector at all is questionable value.

After viewing the table, it is apparent that none the correctors are significantly better than the unc rected form, in fact, all of them are worse than t uncorrected form for at least one of the test integ tions.

On this basis, there is no valid argument for usin corrector formula with the Second Sum Method. T computation time is essentially doubled by the inclusi of the corrector without any gain in accuracy. On t

Solution of $\frac{d^2y}{dx^2} = -y$ *(Testing Truncation Only)*

x	Analytic Solution $\sin x$	Σ^2,Uncorrected 4∇, $\Delta t = .2$	Σ^2, Σ^2 Corrector 4∇, $\Delta t = .2$	Σ^2, Σ Corrector 4∇, $\Delta t = .2$	Σ^2, δ^2 Corrector 4∇, $\Delta t = .2$
5	−.95892427	−.95892*320*	−.95892*433*	−.95892*433*	−.95892*169*
8	.98935825	.98935*612*	.98935*829*	.98935*819*	.98935*512*
14	.99060736	.99060*368*	.99060*758*	.99060*683*	.99060*613*

Solution of $\frac{d^2y}{dx^2} = y$ *(Testing Instability Only)*

x	Analytic Solution e^{-x}	Σ^2 Uncorrected 4∇, $\Delta t = .2$	Σ^2, Σ^2 Corrector 4∇, $\Delta t = .2$	Σ^2, Σ Corrector 4∇, $\Delta t = .2$	Σ^2, δ^2 Corrector 4∇, $\Delta t = .2$
3	.49787068 (−1)	.49798*904* (−1)	.49787*743* (−1)	.49787*235* (−1)	.49785*171* (−1)
5	.67379470 (−2)	.67516*844* (−2)	.67429*677* (−2)	.67392*934* (−2)	.67238*650* (−2)
7	.91188197 (−3)	.10134*813* (−2)	.94898*098* (−3)	.92185*940* (−3)	.80888*140* (−3)

Solution of $\frac{d^4y}{dx^4} = -4y$ *(Test Instability and Truncation)*

x	Analytic Solution $e^{-x}\sin x$	Σ^2, Uncorrected 4∇, $\Delta t = .1$	Σ^2, Σ^2 Corrector 4∇, $\Delta t = .1$	Σ^2, Σ^2 Corrector 4∇, $\Delta t = .1$	Σ^2, δ^2 Corrector 4∇, $\Delta t = .1$
1	.30955987	.30955*989*	.30955*985*	.30955*985*	.30955*979*
3	.70259519 (−2)	.70264*538* (−2)	.70269*610* (−2)	.70270*680* (−2)	.70259*020* (−2)
5	−.64611811 (−2)	−.64627*829* (−2)	−.64632*691* (−2)	−.64639*202* (−2)	−.64547*292* (−2)
8	.33189276 (−3)	.35411*446* (−3)	.35288*331* (−3)	.36416*646* (−3)	.19259*703* (−3)

basis, we must conclude that *the Second Sum Method should be used without a corrector formula* when conditions indicate the method should be used.

Comparison of Integration Techniques

Since the various methods of integration that are being compared rely on different analytical bases, it would be quite difficult, if not impossible, to compare these methods without resort to empirical methods. Simply making reference to the stipulated error terms is inadequate since these estimates are frequently only upper bounds and reveal nothing concerning the stability of the method, i.e. its ability not to magnify the errors that have been made previously.

To compare these methods with reference to their ability to follow accurately the solution of a differential equation where truncation error is most significant, they were all applied to an integration for the sine solution of the second order differential equation

$$\frac{d^2y}{dx^2} = y.$$

Since all solutions of this differential equation (including error modes) remain bounded, the only important source of error in a relatively short integration is truncation. The results of this comparison are summarized in Table I.

To study these same integration methods with respect to their stability, they were applied to the differential equation

TABLE I

Solution of $\frac{d^2y}{dx^2} = y$ *(Testing Truncation Only)*

x	Analytic Solution $\sin x$	Second Sum 7∇, $\Delta x = 1$	Adams 4∇, $\Delta x = .1$	Adams-Bashforth 7∇, $\Delta x = .2$	Milne $\Delta x = .1$	Runge-Kutta $\Delta x = .1$
1	.84147098	.84147098	.84147*169*	—	.84147*120*	.84147*100*
5	−.95892427	−.95892*421*	−.95893*843*	−.95892*435*	−.95892*489*	−.95892*448*
8	.98935825	.98935*822*	.98938*231*	.98935*836*	.98935*801*	.98935*874*
20	.91294525	.91294*535*	.91299*685*	Nonsense	—	.91294*640*

x	Adams-Bashforth 4∇, $\Delta x = .2$
1	—
5	−.95892*849*
8	.98937*139*
20	—

$$\frac{d^2y}{dx^2} = y$$

in an attempt to follow its negative exponential solution. Since this equation also has a positive exponential solution, the integration errors can grow exponentially if the integration technique is potentially an unstable one. The results of this study are given in Table II.

Since most integration problems are not as idealized as either of the examples used in developing Table I and Table II and involve both oscillating functions (which cause large truncation errors) and exponentially growing error terms, a third numerical case was studied.

TABLE II

Solution of $\frac{d^2y}{dx^2} = y$ *(Testing Instability Only)*

x	Analytic Solution e^{-x}	Second Sum 7∇, $\Delta x = .1$	Adams 4∇, $\Delta x = .1$	Adams-Bashforth 7∇, $\Delta x = .2$	Milne $\Delta x = .1$	Runge-Kutta $\Delta x = .4$
3	.49787068 (−1)	.49787*100* (−1)	.49787*057* (−1)	—	.49*795190* (−1)	No entries fo
5	.67379470 (−2)	.6738*1876* (−2)	.6737944*3* (−2)	.6737948*1* (−2)	.6*7984299* (−2)	these precis
7	.91188197 (−3)	.9*1365500* (−3)	.9118814*6* (−3)	.9118821*8* (−3)	.*13588766* (−2)	values of x
11	.16701701 (−4)	.*11349989* (−3)	.1670168*7* (−4)	.1670171*1* (−4)	Nonsense	
13	.22603294 (−5)	Nonsense	.2260327*3* (−5)	.2260332*5* (−5)		
18	.15229980 (−7)		.1522891*6* (−7)	.1522972*4* (−7)		.*13627159* (−7)
34	.17139084 (−14)		.1713678*4* (−14)	.*14011242* (−14)		.*13860855* (−14)
50	.19287498 (−21)		.1928366*9* (−21)	Nonsense		.*14098560* (−21)

	Adams-Bashforth 4∇, $\Delta x = .2$
3	.49787068 (−1)
5	.6737961*8* (−2)
7	.9118858*9* (−3)
11	.1670184*2* (−4)
13	.2260353*5* (−5)
18	.1523022*3* (−7)
34	.1713965*1* (−14)
50	.1928846*3* (−21)

TABLE III

Solution of $\frac{d^4y}{dx^4} = -4y$ *(Testing Instability and Truncation)*

x	Analytic Solution $e^{-x} \sin x$	Second Sum 7∇, $\Delta x = .1$	Adams 4∇, $= \Delta x = .1$	Adams-Bashforth 4∇, $\Delta x = .2$	Adams-Bashforth 7∇, $\Delta x = .2$	Runge-Kutta $\Delta x = .4$
1.0	3.0955987 (−1)	3.0955990 (−1)	3.0955*477* (−1)	3.0956*791* (−1)	—	—
2.0	1.2306002 (−1)	1.2306010 (−1)	1.2306*172* (−1)	1.2307*374* (−1)	1.2306*014* (−1)	—
3.0	7.0259519 (−3)	7.0257*570* (−3)	7.030*1590* (−3)	7.024*3336* (−3)	7.025*8696* (−3)	—
4.0	−1.3861322 (−2)	−1.386*2467* (−2)	−1.385*9552* (−2)	−1.386*7013* (−2)	−1.386*1365* (−2)	−1.380*3029* (−2
5.0	−6.4611811 (−3)	−6.46*31004* (−3)	−6.46*31293* (−3)	−6.46*38852* (−3)	−6.461*0166* (−3)	—
6.0	−6.9260173 (−4)	−6.*8975377* (−4)	−6.*9901350* (−4)	6.*9452092* (−4)	−6.*9238004* (−4)	—
7.0	5.9909419 (−4)	6.*2164638* (−4)	5.*9505955* (−4)	5.*9440127* (−4)	5.*9817605* (−4)	—
8.0	3.3189276 (−4)	3.*7709248* (−4)	3.*6497755* (−4)	3.3098*463* (−4)	3.*2761140* (−4)	3.3*102082* (−4
11.2	−1.3389468 (−5)	−*1.0575953* (−3)	−*8.7215301* (−5)	−*8.3339644* (−6)	*9.1312591* (−5)	−1.*7773536* (−5

This case involved the solutions of the differential equation

$$\frac{d^4y}{dx^4} = -4y$$

which has solutions which are of the form $e^{\pm x}\begin{Bmatrix}\sin\\ \cos\end{Bmatrix} x$, so that both truncation errors and instability problems (due to the presence of positive exponential terms) would be present. The numerical integration attempted to follow the solution $e^{-x} \sin x$ so that instability problems became evident immediately when the numerical solution began to grow at a positive exponential rate. This case should display the interaction between these two major sources of integration error. The results are summarized in Table III.

As can be seen from viewing these tables, neither the Milne predictor-corrector method nor Runge-Kutta could favorable compare with the difference tabl techniques. In order that the Milne Method could ge comparable accuracy to the Second Sum Method, was necessary to use the same integration interval an therefore twice the computing time, and even with th interval, it was more unstable than the not-too-stabl Second Sum Method. On this basis of mediocre ac curacy and great instability, the method was con pletely discarded from further consideration. Th Runge-Kutta method, on the other hand, when co sidered on a basis of equal computing time with th difference table methods (which means four times th interval used for Adams and Second Sum, and twic the interval used for Adams-Bashforth), had simpl atrocious accuracy though it displayed a most interes ing ability to follow a solution without increasing i error. For this reason, *the Runge-Kutta Method is we suited to exploratory computation where accuracy is n important.* In such cases, the ease of starting the integr

tion as well as terminating it at an explicit point, would certainly be of definite value.

A point worth noting in favor of the Runge-Kutta method is that on some integrations of relatively few steps in which the integration interval should vary greatly (as in a space trajectory from earth to moon computed using perturbational methods), the ability to alter the integration interval in an arbitrarily specified manner, rather than simply halving and doubling, may significantly reduce the number of integration steps required. This feature, plus the advantages gained from a self-starting integration procedure, in some cases will indicate that the Runge-Kutta method is the most desirable procedure.

Conclusions

Thus, we arrive at the following general conclusions concerning integration methods to be used for *extensive* integrations in which extreme accuracy is important:

1. Where the integration is known to be stable, the Second Sum Method is the best procedure to employ.
2. Where the integration is known to be unstable, the Adams-Bashforth Method should be used.
3. If an integration is known to be only slightly unstable, the Adams-Bashforth Method should be used with a high number of differences (like seven); however, if the integration is highly unstable, Adams-Bashforth should be used with a relatively small number of differences (like four).
4. If space problems arise within the computer, Adams Method would be a good choice of method as it would involve the shortest program of any of the difference table methods. In this case, no more than four differences should be used since higher differences make this method prohibitively unstable.
5. There is definitely room for better numerical integration methods, particularly for the integration of unstable equations with oscillatory solutions.

Acknowledgement

Important contributions to this paper were made by E. Browne, B. Kubert, and L. Sashkin, all staff members of Aeronutronic Systems Inc.

References

[1] S. Herrick, Step-by-Step Integration of $\ddot{x} = f(x, y, z, t)$ Without a Corrector, *Mathematical Tables and Other Aids to Computation* (*MTAC*), Vol. 5, pp. 61-67, April, 1951.

[2] W. E. Milne, *Numerical Solution of Differential Equations*, John Wiley & Sons, 1953, p. 86.

[3] E. L. Ince, *Ordinary Differential Equations*, Dover, 1956, p. 540.

[4] W. E. Milne, *Numerical Solution of Differential Equations*, John Wiley & Sons, 1953, p. 88.

[5] A. D. Booth, *Numerical Methods*, Butterworths Scientific Publications, 1955, p. 62.

[6] R. W. Hamming, *Stable Predictor-Corrector Methods*, J. Assoc. Computing Machinery, January, 1959, pp. 37–47.

Estimate of the Specific Ionization Caused by Heavy Cosmic Ray Primaries in Tissue or Water*

T. Foelsche†

Abstract

The specific ionization caused by the primary particles of carbon, neon, iron, and niobium along their paths in water is calculated by use of the formulae of Bohr, Bethe and Bloch, for the average stopping power. The maximum specific ionization and shape and also half width of the maximum on the end of the path of the particles is estimated according to the considerations about capture and loss of electrons made by Bohr, Brunings, Knipp, Teller, and Bell.

* From the Air Force Missile Development Center Technical Note AFMDC-TN-59-7 dated April 1959.

† University of Frankfurt-Main, presently on the Scientific and Engineering Staff of the Office of the Chief Scientist, Air Force Missile Development Center, Holloman Air Force Base, New Mexico.

Purpose

The purpose of the following calculation is to estimate the specific ionization caused by heavy cosmic ray primaries along their paths in living tissue or in water, and, provided they do not terminate in disintegration stars, especially at the end of their paths. The main objective of these investigations is to secure an estimate of biological effects. The absolute value of specific ionization at different energies and the shape and the half width of the ionization maximum at the end of the particle's path are of special interest. To correlate the variation of these characteristics with atomic number the primary particles carbon, neon, iron, and niobium are treated as examples.

Theory for Stopping of Heavy Ions in Matter

To find the ionization along the path the main problem involved is calculating the stopping power, or the average energy loss per unit path length (dE/dx) of heavy ions with known velocity v. This energy loss is due to electronic and nuclear collisions with the atoms of the stopping medium. Electronic collisions are interactions of the charged particles with the electrons of the atoms of the stopping medium, while the term "nuclear collisions" refers to recoil phenomena occurring near the end of the particle's path, but not to disintegrations. Having the expression for the stopping power as a function of the velocity or energy per nucleon, the range versus velocity can be found by integration and the specific ionization versus path can be found by eliminating the parameter v or the velocity of the particle.

The Formulae of Bethe-Bloch and Bohr for the Stopping Power [4]

Since the general theory of passage of particles of any atomic weight through absorbers of any atomic weight is very complex, only the formulae and data applicable to the special case of heavy primaries penetrating water are discussed. Because of mathematical difficulties involved in accurate collision theory, semi-empirical data must be used to estimate the stopping power or specific ionization, especially at the very end of the particle's path, or in the low energy range.

The contribution of the electronic collisions, or the electronic stopping power, may be shown in a general expression obtained by Bethe and Bloch.

This equation is:

$$-\frac{dE}{dx} = \frac{4\pi e^4}{mv^2} N(Z_1^{\text{eff}})^2 Z_2 \cdot \ln\left\{\frac{2mv^2}{I} + \psi(1) - R\psi\left(1 + i\frac{Z_1^{\text{eff}} \cdot e^2}{h/2\pi \cdot v}\right)\right\} \quad \text{for } v > Z_2 v_H \tag{1}$$

where:

ψ is the logarithmic derivative of the Γ function,
$R\psi$ denotes the real part of ψ,
$Z_1^{\text{eff}} e$ is the effective charge of incident particle,
v is the velocity of incident particle,
N is the number of stopping atoms per cc,
Z_2 is the nuclear charge of stopping atoms,
I is the average excitation potential of the atoms of the stopping material,
m is electron rest mass, and
v_H is Bohr's electron velocity in the ground state of the hydrogen atom $(e^2/(h/2\pi))$

For $v \gg \frac{Z_1 e^2}{h/2\pi}$ formula (1) is reduced to:

$$-\frac{dE}{dx} = N\frac{4\pi e^4 Z_1^2}{mv^2} Z_2 \left\{\ln\frac{2mv^2}{I}\right\} \tag{2}$$

which is Bethe's formula.

If $v \ll \frac{Z_1 e^2}{h/2\pi}$, since

$$\psi(1) = -0.577 \text{ and } R\psi\left(1 + i\frac{Z_1 e^2}{v \cdot h/2\pi}\right) \approx \ln\frac{Z_1 e^2}{h/2\pi v}$$

we have Bohr's formula

$$-\frac{dE}{dx} = \frac{4\pi e^4}{mv^2} N Z_1^{\text{eff}^2} Z_2 \cdot \ln\left\{\frac{1.123\, mv^3}{Z_1^{\text{eff}} e^2 \cdot I \cdot 2\pi/h}\right\} \tag{3}$$

For relativistic velocities of the incident particle the complete Bethe formula is

$$\frac{dE}{dx} = N\frac{4\pi e^4 Z_1^2}{mv^2} Z_2 \cdot \left\{\ln\frac{2mv^2}{I} - \ln\left(1 - \frac{v^2}{c^2}\right) - \frac{v^2}{v^2}\right\} \tag{2'}$$

In the case of heavy primaries traversing water the Bloch correction of the Bethe formula is significant for the low energy range, because the requirement for the Born approximation:

$$v \gg Z_1 v_H = Z_1 \frac{e^2}{h/2\pi} \tag{1a}$$

is not fulfilled at relatively high velocities of the heavy particle. For the end of the path Bohr's classical approximation has to be used. The condition (1a) is at best valid down to energies somewhat above the lower limits E/A corresponding to $v = Z_1 v_H$ given below:

	$Z_1 v_H$	E/A Mev nucl	Estimated Residual range in H_2O
$_{41}Nb^{93}$	41 v_H	41.3	1000 μ
$_{26}Fe^{56}$	26 v_H	16.6	400 μ
$_{10}Ne^{20}$	10 v_H	2.47	55 μ
$_{6}C^{12}$	6 v_H	0.85	15 μ

Although these residual ranges are less than one per mille of the path of relativistic particles, in this region however, the ionization-maximum occurs.

The Bloch formula (1) can be used for $v > Z_2 v_H$ as shown. That means, considering water as absorber (Z_2 oxygen = 8), about down to $v = 8\, v_H$ corresponding to 1.6 Mev per nucleon. This, for the heavy particles Nb, Fe, Ne, is in the velocity range where the condition $v \ll Z_1 v_H$ is fulfilled because $Z_{\text{water}} < Z_1$, (and, therefore, $v = Z_2 v_H = 8\, v_H < Z_1 v_H$). Hence below this velocity the classical Bohr approximation is valid, which holds independent of the condition $v > Z_2 v_H$. The applicability of the Bloch formula for velocities below $v \approx 1.6$ Mev per nucleon is dubious only for the carbon particle, because in this region $v > Z_2 v_H$ is not fulfilled

ıor is the classical approximation (condition $v \ll Z_1 v_H \approx$ 0.85 Mev per nucleon) valid.

For an estimate of the shape and half width of the specific ionization-maximum and its absolute value, Z_1^{eff} furthermore must be known, and, therefore, the complex mechanism of capture and loss of electrons at low particle velocities must be taken into consideration.

Capture and Loss of Electrons at the End of Path, Z_1^{eff} as Function of v

When passing through matter with velocities high enough so that

$$v \gg Z_1 \frac{e^2}{h/2\pi} = Z_1 v_H ,$$

or when the particle's velocity would be greater than the orbit velocity of its innermost electrons would be, the incident particle is stripped of all its electrons, so that Z_1^{eff} is equal to the nuclear charge of the particle. For decreasing velocity, however, the probability of electron capture by the nucleus increases, whereas the probability of electron loss decreases. Therefore, while slowing down, the particle fills up its electron shells; in short, Z_1^{eff} decreases with decreasing velocity. For particles with higher nuclear charge the electron capture process begins at much higher velocities than with very light particles like protons and alpha particles and, therefore, for heavy primaries the decrease of charge cannot be disregarded for obtaining the stopping power as a function of energy or a range energy relationship of any correctness. A first approximation of the quantitative relation between effective charge and velocity of the particle, based on the Fermi-Thomas statistical model of the atom was given by Bohr, who assumed that all electrons having a lower orbital velocity than the particle velocity v would be shaken off, and he gave for the effective charge of the particles:

$$Z_1^{\text{eff}} = Z_1 \frac{1}{3} \frac{v}{v_H}$$

above, $v_H = e^2/(h/2\pi)$ is the Bohr velocity of the electron of the H atom in the unexcited state, Z_1 being the atomic number of the particle. Knipp and Teller [2] and Brunings, Knipp and Teller [2a] gave a variation of charge with velocity, which is somewhat closer to experimental evidence, setting the critical electron velocity v_e (all electrons with $v_{(\text{orbital})} > v_e$ remaining in the ion) *proportional* to the particle velocity v, introducing the parameter γ by the equation

$$v_e = \gamma v$$

Using the Fermi-Thomas model of the absorber atom they calculated the effective charge as a function of velocity assuming:

1. That the characteristic velocity v_e is equal to the root-mean-square velocity of the *energetically most loosely* bound electron or, alternatively,
2. equal to the root-mean-square velocity of the *outermost* electron of the particle.

The parameter γ was determined from the best experimental evidence available, after computing the stopping power, as a function of velocity, for different parameters γ, chosen within certain limits. The variation of effective charge with v_e was given in graphical form [2a], p. 658. With the knowledge of the parameter γ this relationship can be converted into a relation between particle velocity and Z_1^{eff}.

C. I. Bell [3] has published a more detailed theory about the capture and loss of electrons of fast ions. He obtains values for Z_1^{eff} at a certain velocity about 10 per cent smaller than those of Brunings-Knipp-Teller's second assumption. In order to give order of magnitude calculations in closer agreement with the more recent theoretical treatments of the subject of charge exchange of fast ions in matter, the second assumption of Brunings, Knipp, and Teller and the Bohr formula [3] for calculation of ionization versus energy is used. The relations between Z_1^{eff} and v_e which are used are represented in Fig. 1. Substituting Z_1^{eff} into Bohr's formula, [3], we can calculate the energy loss that is due to *electronic collisions* in the low energy region.

Nuclear Collisions

The capture and loss of electrons in the slowing down process has a marked effect on the mechanism for the energy loss of the particle toward the very end of its path. In this lowest energy region, collisions between the particle and the nuclei of the stopping material are more important than the electronic excitation processes. An estimate using the formulas and considerations of Bohr [4] shows that the contribution of the nuclear stopping power measured in per cent of the electronic stopping power is:

at 0.3 Mev per nucleon: Nb $\backsim$ 4%, Fe $\backsim$ 2%, Ne, C < 0.5%

at 0.1 Mev per nucleon: Nb > 100%, Fe $\backsim$ 50%, Ne $\backsim$ 8%, C $\backsim$ 3%

We are at 0.3 Mev per nucleon energy in that region where the ionization versus path steeply decreases. For the rough first estimate of the half width given here it is considered as sufficient to calculate the specific ionization versus path down to 0.2 Mev per nucleon, taking into consideration only electronic collisions. The given numbers show that above this velocity the contribution of nuclear collisions to the overall energy loss per unit path can be neglected in the first approximation. For the residual range below 0.2 Mev per nucleon empirical data are used.

Approximative Calculation of Specific Ionization Along the Path

If the stopping power dE/dx, the decrease of the kinetic energy per unit path of the primary of the atomic weight A, is given as function of the particle's kinetic energy E, the range R is obtained by:

$$R = \int_v^0 dx = \int_v^0 \frac{dE}{\left(\frac{dE}{dx}\right)}$$

$$= A \int_v^0 \frac{v\,dv}{\left\{\frac{dE(v)}{dx}\right\}} = A \int_{E/A}^0 \frac{dE/A}{\left\{\frac{dE}{dx}\left(\frac{E}{A}\right)\right\}}$$

Eliminating the parameter E/A or v from dE/dx and R, or plotting dE/dx as a function of the range R, we obtain the stopping power along the path. Taking into consideration that for each ion pair formed in water 35 ev is consumed, we get the number of ion pairs formed by the primary per unit path or the specific ionization along the path dividing dE/dx by 35 ev (if dE/dx is given in ev per unit length). These 35 ev comprise not only the energy for ionization of the stopping atoms, by slow and fast secondary electrons, but also the comparatively large amount of energy, which is lost because of mere excitation and dissociation of the molecules or atoms.

The basic curves of energy loss dE/dx or specific ionization as function of v or E/A (Fig. 2) for the heavy ions of Nb, Fe, Ne, C, are constructed from two parts. Above $v \approx Z_1\, e^2/(h/2\pi)$, dE/dx is given by the Bethe formulae [2] and [2']. For such high energies the ratio of the stopping powers of two particles having different nuclear charge Z_1, Z_1', but travelling with the same velocity through the same material is:

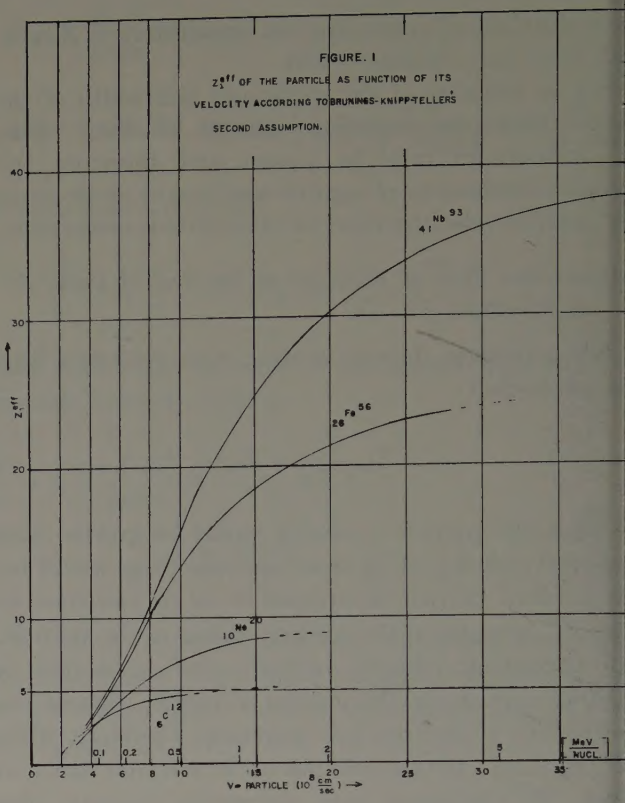
$$\frac{dE/dx}{(dE/dx)'} = \frac{(Z_1)^2}{(Z_1')^2}$$

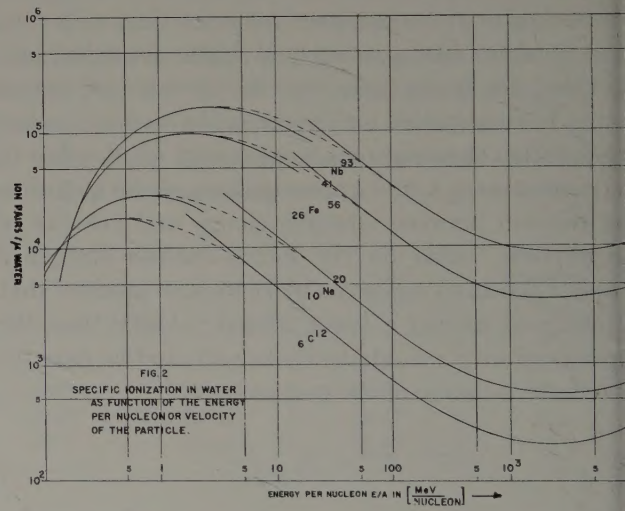
For high velocities the stopping power of any charged particle can be derived from that of the *proton* in the same medium, which is well known. In Fig. 2 the values for protons in water given in [5] multiplied with Z_1^2 are used. In the same manner, by comparison with the travelled distance of a proton, the range difference between the high initial velocity v and the lower limit $v_1 \gg Z_1 v_H$ of the nucleus Z_1 can be calculated by the integral

$$\Delta R = \int_v^{v_1} \frac{dE}{\left(\frac{dE}{dx}\right)} = \frac{1}{Z_1^2} \cdot \int_v^{v_1} \frac{dE}{\left(\frac{dE}{dx}\right)_{\text{proton}}}$$

or equivalently in terms of the kinetic energy per nucleon E/A, when A is the mass number of the particle's nucleus,

$$\Delta R_{Z_1, E/A} = \frac{A}{Z_1^2}\, \Delta R_{\text{prot}, E/A}$$

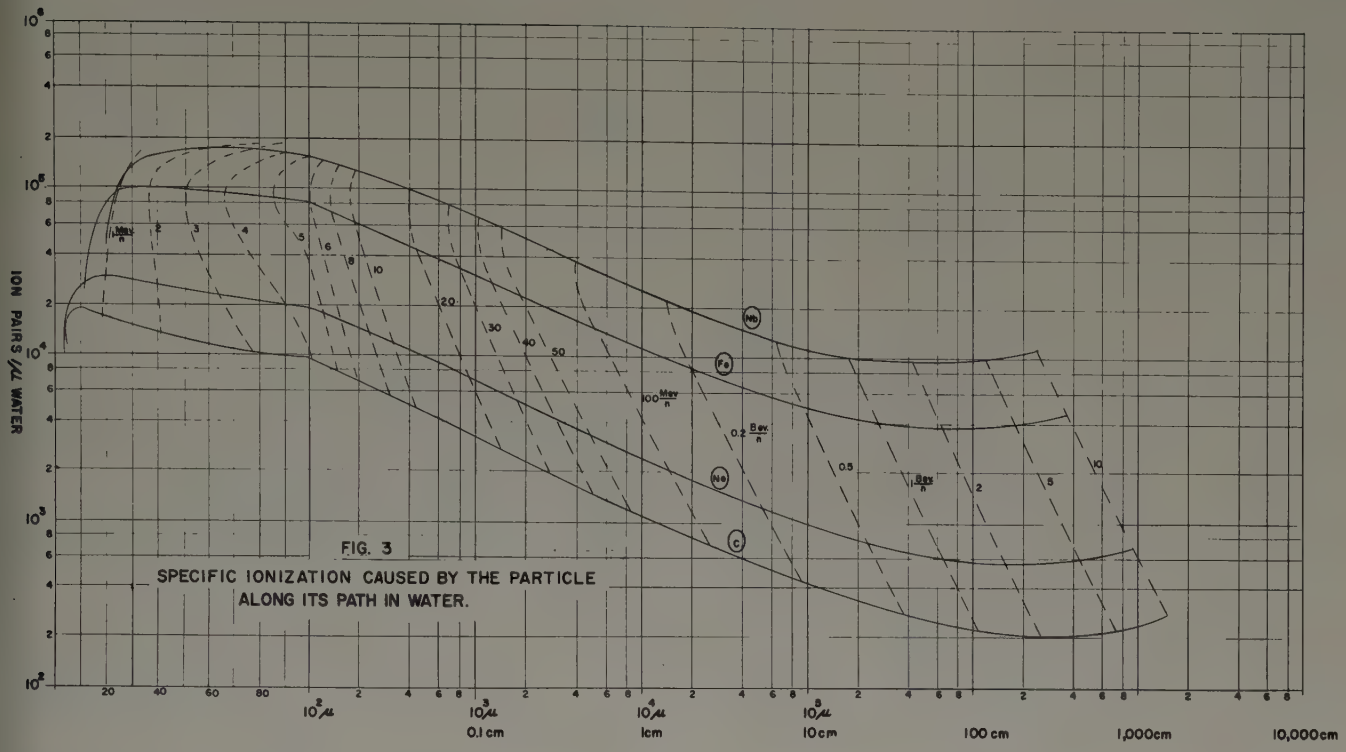




For low energies of the particle the Bohr formula for the stopping power is used, which in terms of energy per nucleon E/A reads for water ($Z_2 = 10$, $I = 66$ e

$$\frac{dE}{dx} = 79.4\, \frac{(Z_1^{\text{eff}})^2}{E/A} \ln\left(\frac{120(E/A)^{3/2}}{Z_1^{\text{eff}}}\right) \left[\frac{\text{Mev}}{\text{cm}}\right]$$

Z_1^{eff} as a function of v or E/A is inserted according the second assumption of Brunings, Knipp and Tel



(Fig. 1) with the following values for γ:

	γ
Nb	0.37
Fe	0.41
Ne	0.45
C	0.60

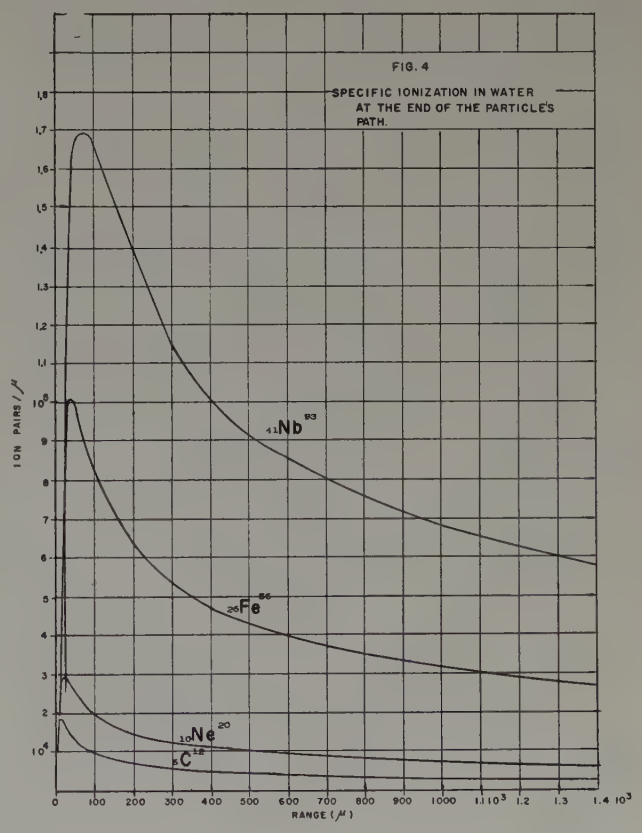
For the intermediate energy region in this first approximation the Bloch formula for the stopping power is interpolated by the dashed lines which fit in the exact values at the lower and higher energies (Fig. 2). In this way dE/dx is obtained down to energies 0.2 Mev per nucleon. The travelled distance, or range difference

$$\Delta R = \int_{v_0}^{v} \frac{dE}{\left(-\frac{dE}{dx}\right)} = A \cdot \int_{0.2\left[\frac{\text{Mev}}{\text{nucl}}\right]}^{E/A} \frac{dE/A}{\left(-\frac{dE}{dx}\right)}$$

between a given initial velocity v and the lower limit v_0 corresponding to 0.2 Mev per nucleon as function of the variable velocity v are then calculated by integration with an analog computer. To obtain the total range, we have to add the empirical residual ranges at 0.2 Mev per nucleon which amounts, for the different particles, to:

	Residual Range at 0.2 Mev per nucleon
Nb	20 μ
Fe	12 μ
Ne	5 μ
C	4 μ

For Nb the value for the light fission fragments is used. The value of Fe is interpolated because no experimental data were available. The values for Nb, Ne, C, are taken from Brunings, Knipp and Teller [2a]. From the summation the total range-energy relations listed in Table I result.



Plotting the stopping power or the specific ionization of Fig. 2 at a given energy per nucleon versus the total range at the same energy per nucleon the specific ionization along the path is obtained (see Fig. 3 and 4).

Results

Figure 3 shows the specific ionization along the total path of relativistic particles having initial velocities of 10 Bev per nucleon. The maximum of the ionization at the end of the path is asymmetrical. The ionization decreases very steeply at the very end of the path (Fig. 3 and 4). Even for the heavy primary Nb this steep decrease takes place in a residual range of not more than $\approx 30\ \mu$. In comparison with the width of this descending part the half width of the maximum is large ($\approx 550\ \mu$ for Nb). The ascent begins slowly even if the maximum is a sharp peak. Its half width does not exceed the last $\mu\mu$ of the particle's path as is shown in Fig. 4.

To compare the biological effects of heavy primaries with the effect of alpha particles, it is of interest to know along which part of the primary's track the specific ionization is higher than the maximum specific ionization caused by an alpha particle (max. 10^4 ion pairs per micron, half width $\approx 10\ \mu$). Those features of the ionization curve, as the absolute values of the maxima, its half widths, and those parts of the path where the ionization is higher than 10^4 ion pairs per micron and higher than 10^5 ion pairs per micron are listed in Table II.

The half width of the maximum increases with atomic weight (not uniformly) by the factor 6 if the atomic weight increases by the factor 8. The increase of the ionization-maximum itself is roughly proportional to the atomic weight of the incident particle.

It must be emphasized that it is the *absolute* value of the maximum which heavily depends on the relation between Z_1^{eff} and v. Using Brunings, Knipp and Teller's first assumption with empirical values of γ the calculation results in 40 per cent higher maxima. The above given values are to be regarded as rough estimates.

The main purpose in making these calculations is to have a survey of the ranges of heavy primaries in tissue and to identify that part of the particle's path along which the particles begin heavily to ionize.

To give an example, in Fig. 3 it is shown that an Fe primary having the relatively low velocity corresponding to 0.5 Bev per nucleon (28 Bev of energy) has a range of about 9 cm in tissue and causes a specific ionization higher than 0.8×10^4 ion pairs per μ along the last 2 cm of its track. In comparison with this, a 4 Mev alpha particle, as it is emitted from naturally radioactive substances, has a range of about 30 μ in tissue. The effect of the heavy primary in the last 2 cm therefore corresponds at least to the effect of 670 such alpha particles.

Actually the specific ionization along this part of the track is up to 10 times higher and is distributed over a broader volume [6], so that the biological effect should be considerably higher.

TABLE I

Total Average Range in μ (cm)

Mev/n	0.2	0.6	1	2	3	4	5	6	8	10
	(μ)									
Nb	20	25	30	65	80	95	110	125	155	195
Fe	12	19	24	37	52	67	87	104	140	190
Ne	5	15	21	40	65	92	125	161	241	335
C	4	10	18	41	71	109	150	196	306	440

Mev/n	20	50	100	200	500	1000	2000	5000	10^4
				(cm)					
Nb	405	1450	3925	1.41	6.42	17.9	43.7	123	247
Fe	455	1860	5300	2.05	9.42	26.8	65.3	184	370
			(cm)						
Ne	921	5371	1.54	5.21	23.0	65.0	158	445	894
C	1440	8855	2.55	8.67	38.3	108	263	742	1490

TABLE II

Element	Maximum				Part of the path with ionization where $\frac{\text{ion } p}{\mu}$ is:	
	Energy Mev/n	Residual Range	Ion Pairs (μ)	Half Width	$>10^4$	$>10^5$
$_6C^{12}$	0.5	9 μ	1.9×10^4	95 μ	85 μ	—
$_{10}Ne^{20}$	0.7	20 μ	3.0×10^4	195 μ	500 μ	—
$_{26}Fe^{56}$	1.5	38 μ	10^5	340 μ	1.3 cm	20 μ
$_{41}Nb^{93}$	2.0	70 μ	1.7×10^5	550 μ	>100 cm*	375 μ

* practically.

Acknowledgement

The author expresses his sincere appreciation to Mr. W. F. Haldeman, aeronautical engineer, and Mr. James V. West, physicist, at the Air Force Missile Development Center, for the necessary computations with analog computer, and to Horst Foelsche for numerical calculations.

References

[1.] N. Bohr, *Phys. Rev.*, Vol. 58, p. 654, 1940.

[1a.] N. Bohr, *Phys. Rev.*, Vol. 59, p. 270, 1941.

[2.] F. Knipp, and E. Teller, *Phys. Rev.*, Vol. 59, p. 659, 1941.

[2a.] F. M. M. Brunings, F. K. Knipp, and E. Teller, *Phys. Rev.*, Vol. 60, p. 657, 1941.

[3.] C. I. Bell, *Phys. Rev.*, Vol. 90, p. 548, 1953.

[4.] Bethe-Ashkins, "Passage of Radiation Through Matter," E. Segré, ed., *Experimental Nuclear Physics*, Vol. I, p. 168, John Wiley and Sons, Inc., 1953.

[5.] Hine Bromwell, *Radiation Dosimetry*, p. 226, Academic Press, Inc., 1956.

[6.] Hermann Schaefer, Naval School of Aviation Medicine Research Report *NM 001 059.13.04*, 5 August 1952.

Some Aspects of the Electrical Properties of the Upper Atmosphere*

P. A. Goldberg†

Abstract

Evaluations are presented of electrical properties of the D-region of the ionosphere for altitudes of 150,000 to 300,000 feet. Certain unique regimes of electrical behaviour of this portion of the ionosphere are pointed out. Covered are altitude profiles of a.c. electrical conductivity, dielectric constant, index of refraction, vertical attenuation factor and penetration depth, and long oblique path radio signal attenuation for frequencies from d.c. to 100 megacycles. Applications to spacecraft environment problems are treated.

Introduction

The altitude of about 65 kilometers marks an important transition region for the geophysical system and its environment. Below this altitude lies more than 99 per cent of the earth's "atmosphere". Above it is the beginning of the "electrical sea" which surrounds our planet, reaching an ionization maximum at some 300 km and then extending diffusely throughout extra-terrestrial space.

The study presented here is a preliminary analytical model of the electrical properties of the beginning or "surface" region of this electrical sea, the D-region of the ionosphere. A principal motivation in this study is the consideration of the electrical environment of spacecraft. Such studies previously [1], [2] have concentrated on the local electrical environment due to ionization generated in shock waves. This study is concerned with the properties of the ambient electrical environment due to ionization in the D-region. Other applications of the work are connected with the properties of the terrestrial waveguide such as appear in problems of Very Low Frequency propagation [3], Whistler propagation [4], [5], and Super Low Frequency propagation [6], [7].

Figure 1 places in perspective the role of the D-region in spacecraft environment problems. The figure shows the principal ionosphere regions in terms of the vertical distribution of electron density. We see that all classes of spacecraft can be subjected to the influence of the D-region. Hypersonic gliders may operate just below, in or above the D-region. High altitude research rockets and long range missiles pass through it. Satellites and lunar and planetary probes are above it for most of their operations. Thus, all such craft generally come directly in contact with the D-region or must "look" through it.

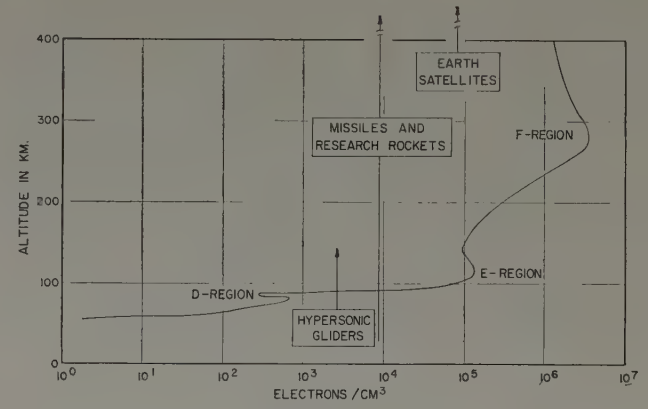


FIG. 1. Typical distribution of free electrons in the upper atmosphere. Proposed hypersonic gliders will operate in the lower ionosphere. Missiles, rockets, and satellites pass through or must "see" through this lower, electrically lossy portion of the ionosphere.

The D-region has certain properties quite different from the other ionosphere regions. Although the region has the lowest electron density it has the highest electron collision frequency of all the ionosphere regions. This makes for marked electrical dissipative effects in certain electrical regimes. In other regimes the region has a pronounced "metallic" type of behaviour. A specific objective of this study is the pointing up and evaluation of such prominent features of the D-region.

More generally, the objective of this analysis is to develop a broad picture of lower ionosphere electrical properties for a wide range of frequencies. The analysis proceeds along five major steps:

1. A model of the electronic parameters; the electron density and collision frequency, is set up.
2. The variation with altitude of the conductivity and dielectric constant is evaluated for a wide range of frequencies.
3. The propagation parameters, the index of refraction and the attenuation factor are then evaluated.
4. The latter results are examined for the two important cases of long vertical and horizontal propagation paths in the D-region.
5. The ranges of magnitude of all these basic properties are then summarized to give an over-all electrical picture of the D-region.

* Presented at the Fifth Annual Meeting of the AAS, Washington, D. C., Dec. 1958.

† Rand Corporation, Santa Monica, Calif.

The analysis covers the wide frequency range of 10 to 10^8 c/s. To make the problem tractable for the wide range of variables thereby encountered, certain limiting considerations are applied. The analysis is carried out for the case of a day-time, summer, electron density model. This is a maximum ionization condition. In evaluating the electrical and propagation parameters the earth's magnetic field is neglected. The results are nevertheless good for the cases where the electron gyro frequency is not important compared to the collision and signal frequencies. Also, and more generally, when the electric vector of the signal is parallel to the geomagnetic vector the analysis holds. In the propagation effects analyses the signal wavelength is assumed small compared to distance changes of local refractive index. This ray treatment is valid for the higher frequencies. For the lower frequencies such as 10 kc the analyses still show the trend of effects and provide results directly applicable to techniques developed for handling such cases, like that by Poeverlein [8].

The D-Region Electron Parameters Model

The two basic electronic factors determining the electrical characteristics of the D-region are the number of electrons per unit volume, N, and their frequency of collision, ν, with air molecules and atoms. These parameters have considerable variation in time. The D-region ionization is a maximum near mid-day and is present only in traces at night. Also, in addition to a seasonal variation there is a strong control of electron density by solar activity such as flares which produce radio blackouts by enhancement of D-region ionization.

The D-region electrical model developed here is based on a "typical" representation of the electron density. For the vertical distribution of electron density we use the daytime summer noon model of Nertney [9]. The electron collision frequency model used here is that of Nicolet [10]. Our resultant electromagnetic model will therefore represent maximum electrical effects for a geophysically quiet day. The electron models used are shown in Fig. 2.

Ionosphere Electrical Parameters—General

In developing the picture of the D-region electromagnetic properties we first construct models of the conduction and dielectric properties. From these we then construct wave propagation models. The formulae to be used in these constructions are taken from ionospheric theory. For the background and development of such theory the reader is referred to the brief account by Stratton [11] or the comprehensive review by Mitra [12].

Because there is some variation in convention and usage of terminology we shall first set down some simple definitions. The differences of terminology have to do with inter-relations of conduction, convection

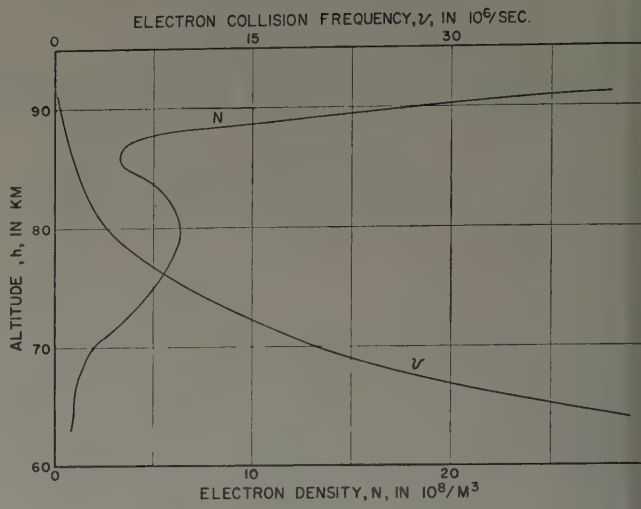


FIG. 2. Typical D-Region altitude profiles of electronic ga parameters. The electron density is relatively small and th collision frequency is relatively quite high. Nevertheles both parameters can contribute either strongly or weakly electrical effects depending on the electrical excitatio frequency.

and displacement currents which lead to variatio in use of terms such as "complex conductivity" "complex dielectric constant".

Our definitions are based on Maxwell's equation f currents

$$\nabla \times H = i_{cd} + i_{cv} + i_d$$

where H is the magnetic intensity, and i_{cd}, i_{cv} and are conduction, convection, and displacement current respectively. Then, for an electric field, E, with angul frequency, ω, and the ohmic condition, $\sigma \neq \sigma(E)$,

$$\nabla \times H = \sigma E + j\omega\epsilon_0\epsilon_r E$$

where σ is the conductivity, ϵ_r the permittivity (dielectric constant) relative to that of free space, (all units in this paper are those of the rationalize m.k.s. system). Contributions for all currents whic appear in the real number, σ, in the "in-phase" membe σE, will be lumped together and called the "electric conductivity". Contributions to the $\pm 90°$ out-o phase member appearing collectively in the real numb ϵ_r will be called the dielectric constant or relative pe mittivity. With these definitions, in the expressio below, the conductivity is always positive; the d electric constant may be either positive or negati and is always less than unity.

The local conductivity of an ionospheric region given by the expression,

$$\sigma = (Ne^2\nu)/m(\omega^2 + \nu^2) = (\epsilon_0\omega_c^2\nu)/(\omega^2 + \nu^2)$$

where e and m are the electronic charge and mas respectively. In the alternative form, $\omega_c^2 \equiv (Ne^2)/(me$ is the "characteristic" or "plasma" frequency.

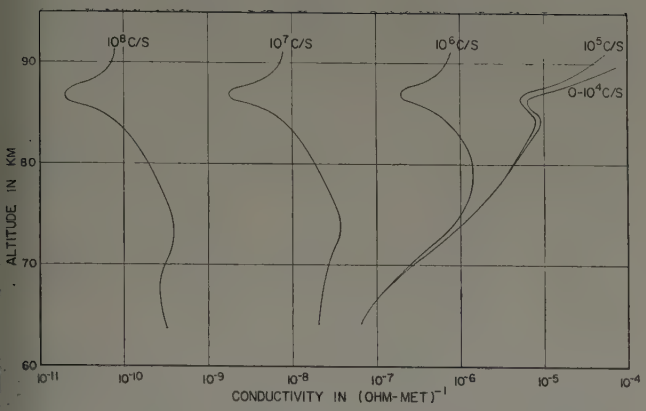


FIG. 3. Electrical conductivity of the lower ionosphere for altitudes from 65 to 91 km and for frequencies from DC to 100 mc. Although numerically small, the cumulative effects of conductivity can lead to large dissipation of electrical energy up to frequencies of 10 mc, and even higher frequencies on geophysically active days.

The dielectric constant is given by the expression

$$\epsilon_r = 1 - (Ne^2)/(m\epsilon_0)(\omega^2 + \nu^2) = 1 - \omega_c^2/(\omega^2 + \nu^2). \quad (2)$$

The above expressions assume that effects of any ambient magnetic field can be neglected. Remarks on magnetic field effects will be deferred until the later section on propagation.

D-Region Electrical Properties

Electrical Conductivity

The variation with altitude and frequency of D-region conductivity is shown in Fig. 3. These values were obtained by applying expression (1) directly to the electronic model of Fig. 2. Additional values for frequencies higher than the 10^8 c/s figures shown can be generated by simply translating the 10^8 c/s curves two decades to the left for each decade change in frequency. This is because at this end of the frequency regime the conductivity is to excellent approximation given by $Ne^2\nu/(m\omega^2)$. At the lowest frequency, $\sigma = Ne^2/(m\nu)$, and the structure of the conductivity profile represents principally the effect of the falling off of ν with altitude. (The collision frequency changes by a factor of a hundred for only a twentyfold change in electron density.)

Comparing high and low frequency cases, at the highest frequencies, the conductivity $\sigma_{hi} \sim N\nu/\omega^2$ and at the lowest frequencies $\sigma_{lo} \sim N/\nu$, and so $\sigma_{hi}/\sigma_{lo} \sim \nu^2/\omega^2$. Therefore the conductivity at high frequencies is much lower than that at low frequencies and also does not change with attitude much compared to the low frequency case.

A comparison of D-region conductivity with other ionized gas conditions and conducting substances is made in Table 1. It is seen that the D-region is much less conducting than metals, ionized shock waves, and the higher regions of the ionosphere. (The shock wave values are estimates based on classical kinetic theory

TABLE 1

Comparison of Electrical Conductivities

Material	D.C. Conductivity in (ohm-met)$^{-1}$
Copper (20°C)	6×10^7
Carbon (2500°C)	1×10^5
Carbon (0°C)	3×10^4
Tellurium	5×10^2
Shock wave (Mach 20, 56 km)	4×10^2
Shock wave (Mach 15, 56 km)	1×10^2
Shock wave (Mach 13, 56 km)	2×10^1
Manganese dioxide	2×10^1
Ionosphere, F-region (300 km)	2×10^1
Sea water	5×10^0
Ionosphere, E-region (110 km)	2×10^{-2}
Clay	1×10^{-2}
Distilled water (25°C)	1×10^{-4}
Ionosphere, D-region (90 km)	1×10^{-4}
Ionosphere, D-region (80 km)	4×10^{-6}
Slate	1×10^{-6}
Ionosphere, D-region (70 km)	3×10^{-7}
Mica	5×10^{-16}
Fused quartz	2×10^{-17}

molecular diameters [2].) Nevertheless, as shown later although the local conductivity is relatively low the cumulative effect for paths through this region are quite large for a wide range of field frequencies. In addition it will be seen the D-region has a strong metallic characteristic at low frequencies in the sense that it can have a very high index of refraction and reflectivity.

Dielectric Constant

The dielectric constant model shown in Fig. 4 is generated by applying expression (2) to Fig. 2. For frequencies above 10^8 c/s the D-region dielectric constant is essentially unity and dielectric effects are negligible. This is due to the effect of electron inertia at these frequencies and the resulting negligible displacement currents. At the other extreme, at suffi-

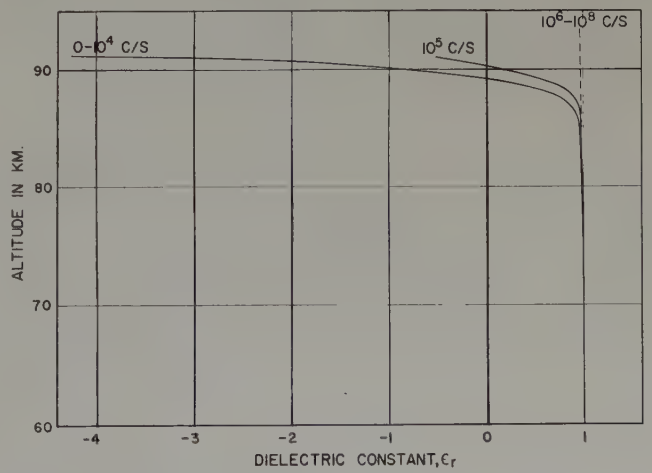


FIG. 4. Altitude profile of the dielectric constant of the lower ionosphere for frequencies from DC to 100 mc.

TABLE 2

Comparison of Dielectric Constants

Material	Dielectric Constant at:		
	10^2 c/s	10^5 c/s	10^8 c/s
Ionosphere, D-region (90 km)	−0.88	0.19	1.0
Ionosphere, D-region (80 km)	0.98	0.99	1.0
Ionosphere, D-region (70 km)	0.99	0.99	1.0
Air (0°C, 1 atm)		1.0	
Air (11°C, 180 atm)		1.1	
Balsa wood	1.4	1.4	1.3
Vaseline	2.2	2.2	2.2
Polystyrene	2.6	2.6	2.6
Ice (−12°C)		4.2	3.5
Neoprene	7.5	5.9	3.4
Ethyl alcohol		25	23
Barium oxide		34	
Water (85°C)		58	57
Barium titanate		4400	

ciently low frequencies, the dielectric constant attains large negative values. This is again because of electron inertia, this time, however, appearing in the form of displacement currents lagging the exciting field.

In Table 2 the dielectric constants of the D-region are compared with some other dielectrics. The table points up the unique dielectric feature of a plasma of ϵ_r being less than unity and possibly negative. As has been pointed out elsewhere [13] this feature of the ionosphere may attain importance in the operation of antennas. Because the system radiates into a region of modified wave impedance, antenna detuning can occur. For the relatively low altitude of the D-region, local ionization formed in the hypersonic shock waves of a spacecraft can mask D-region "contact" or local effects. The ambient D-region local effect can still be important, however, for locations on the spacecraft in expansion flow and for thin extended antennas about which slip or free flow is occurring.

Ionosphere Propagation Parameters—General

The effect of the D-region on the propagation of electromagnetic fields will be developed below in terms of the index of refraction, n, and index of absorption, k. These factors are related to σ and ϵ_r as

$$(n - jk)^2 = \epsilon_r - j\sigma/(\omega\epsilon_0)$$

or explicitly as

$$\begin{aligned} 2k^2 &= -\epsilon_r + [\epsilon_r^2 + (\sigma/\omega\epsilon_0)^2]^{1/2} \\ &= -[1 - \omega_c^2/(\omega^2 + \nu^2)] \\ &\quad + \{[1 - \omega_c^2/(\omega^2 + \nu^2)]^2 + (\sigma/\omega\epsilon_0)^2\}^{1/2} \end{aligned} \tag{3}$$

and

$$\begin{aligned} 2n^2 &= \epsilon_r + [\epsilon_r^2 + (\sigma/\omega\epsilon_0)^2]^{1/2} \\ &= [1 - \omega_c^2/(\omega^2 + \nu^2)] \\ &\quad + \{1 - \omega_c^2/(\omega^2 + \nu^2)^2 + (\sigma/\omega\epsilon_0)^2\}^{1/2}. \end{aligned} \tag{4}$$

The full magneto-ionic expressions for the indexe as developed by Appleton and Hartree [14] contai terms involving electron gyrofrequency componen transverse and longitudinal to the direction of propa gation of a plane wave front. The expressions (3) an (4) represent the cases where these gyrofrequenc terms can be neglected such as for the lower regions the D-region where the collision frequency effect dominant or when the gyro frequency component small because of wave orientation.

It should be noted that both k and n are alway positive although they can become quite small. Thes propagation properties and the others evaluate in the following sections were obtained by applyin expressions (3) and (4) to the electrical models Figs. 3 and 4.

D-Region Propagation Properties

Index of Refraction

The altitude variation of the D-region refractiv index for frequencies to 10^8 c/s is shown in Fig. We see that in the low frequency regime between 1 to 10^3 c/s the D-region refractive index becomes ver large, attaining values as great as 400. Because th

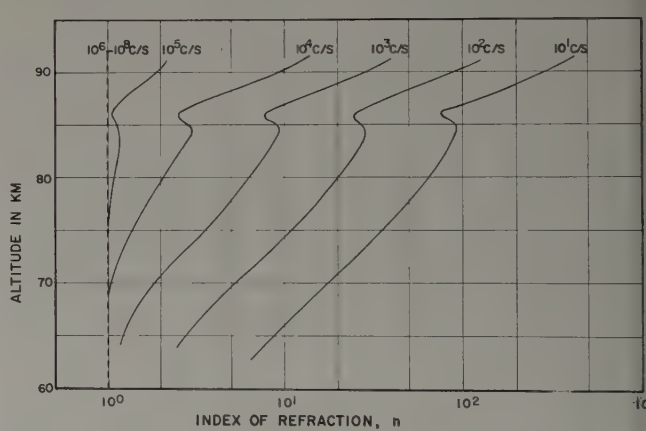


FIG. 5. Altitude profile of D-Region index of refraction fo frequencies from 10 to 10^8 c/s. At low frequencies such as 1 c/s, the lower ionosphere is strongly reflecting because of th large values and gradients of refractive index. This makes th earth-ionosphere system a very good waveguide at suc frequencies.

absorption index (discussed in the next section) relatively low, the D-region at these frequencies is thu very "shiny" or highly reflecting. This state of affai is similar to free electron gas conditions in metals an so the D-region may therefore be appropriately de scribed as "metallic" at these frequencies.

At higher frequencies, the refractive index falls t lower values, but does not fall to less than two pe cent of unity. It should be noted, however, that a higher altitudes such as in the E and F-regions whe the collision frequency becomes very small, the inde of refraction can approach (but never becomes les than) the value zero. This latter effect occurs fo

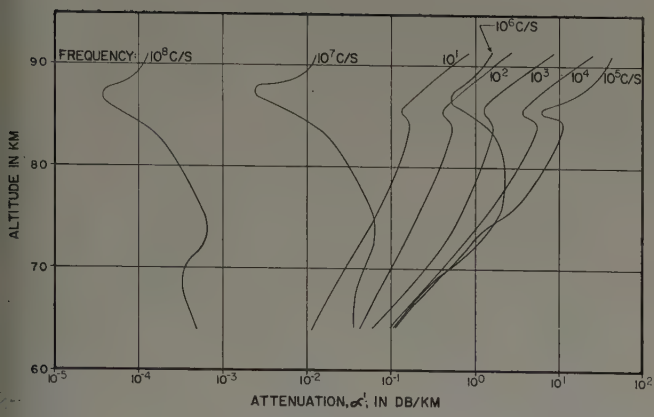


FIG. 6. D-Region altitude profile of local signal attenuation for frequencies from 10^1 to 10^8 c/s. The transition curve of 10^6 c/s corresponds to the approximate condition of maximum energy transfer from the applied electromagnetic field to neutral gas particles in the ionosphere. Energy transfer is most effective where the excitation frequency approximates the local electron collision frequency.

excitation frequencies slightly greater or slightly less than the plasma frequency, as indicated in expression (4).

Attenuation Factor

The local electrical absorption properties of the D-region are shown in Fig. 6. For convenience in propagation calculations the values are given in terms of α', the attenuation in decibels per kilometer instead of absorption index k, where $\alpha' = 8680\ k\omega/c$, and c is the free space velocity of light.

At low frequencies like 10^4 c/s the medium is only moderately absorbing. Thus, such waves do not penetrate very deeply into the D-region. Coupled with large refractive index and index gradients this causes the medium to be an excellent reflector at such frequencies.

At frequencies of 10^5 to 10^6 c/s the excitation frequency is approximately equal to the electron collision frequency. Maximum coupling for energy transfer into conduction heating then occurs and the absorption becomes a maximum.

At frequencies greater than 10^6 c/s the absorption decreases rapidly. Nevertheless, even for, say, the case of 10^7 c/s, where α' is only about 5×10^{-2} *db/km*, certain types of cumulative effects can make absorption at high frequencies significant.

Effects of D-Region on Propagation over Long Paths

Cumulative effects of D-region attenuation over long paths are considered in the following two sections. Among the problems involved for such long path geometries are: communications between spacecraft both through and above the D-region; radiation of electrical noise from the earth into extra-terrestrial space and vice versa; signalling between spacecraft and ground stations. A particular problem in regard to radiation is that of whistler propagation (1 kc–20 kc) via magneto-ionic ducting into extra-terrestrial space. The evaluations developed here apply directly to such propagation since for the lower D-region the collision frequencies are much higher than the magnetic gyro frequencies of the electrons.

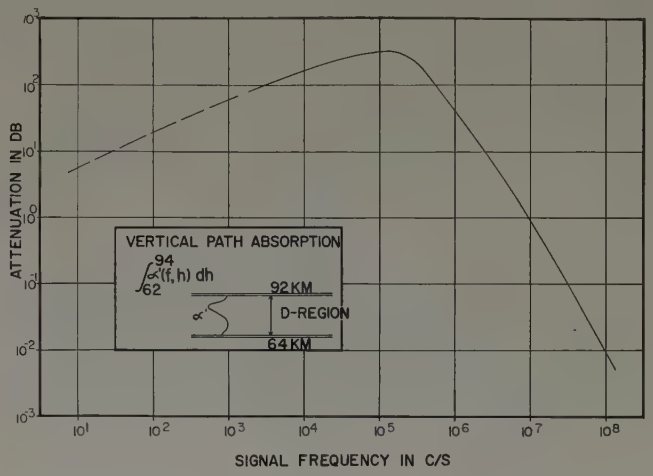


FIG. 7. Total absorption of signals for vertical direction through the D-Region for frequencies from 10 to 10^8 c/s. Such cumulative effects of energy dissipation can lead to signal losses as great as 300 *db* at 10^5 c/s. Most natural and artificial signals originating on the earth below frequencies of 5 mc. should be difficult to observe on satellites because of this filtering action of the D-Region. Special exceptions are ionospheric leakages such as by the proposed whistler ducts.

Vertical Path Absorption

The total absorption experienced by signals passing vertically through the D-region is shown in Fig. 7. The values shown are the integrated values of the attenuation factor:

$$\alpha(f)_{\text{total}} = \int_{62}^{94} \alpha'(f, h)\ dh.$$

The values for the low frequencies are shown dotted to emphasize that these values represent only absorptive losses calculated from α'. There will be, in addition, at these frequencies, effects due to loss by scattering because of the high reflectivity of the region. At frequencies greater than 10^7 c/s the losses become negligible (less than one *db*).

Horizontal Path Absorption

Total absorption for horizontal paths 1000 *km* long are shown in Fig. 8. The values are calculated on the basis of a plane-stratified D-region. This class of path is characteristic of spacecraft orientations such as: two satellites communicating along a path which nearly grazes the earth or is reflected from the earth's surface near grazing incidence. Another orientation in this class would be that of two hypersonic gliders "cruising" in the lower ionosphere.

It is seen that now signals at 10^7 c/s could experience losses of the order of 50 *db*. It is not until frequencies as high as 10^8 c/s are used that losses can fall to less than one *db* (on a geophysically quiet day).

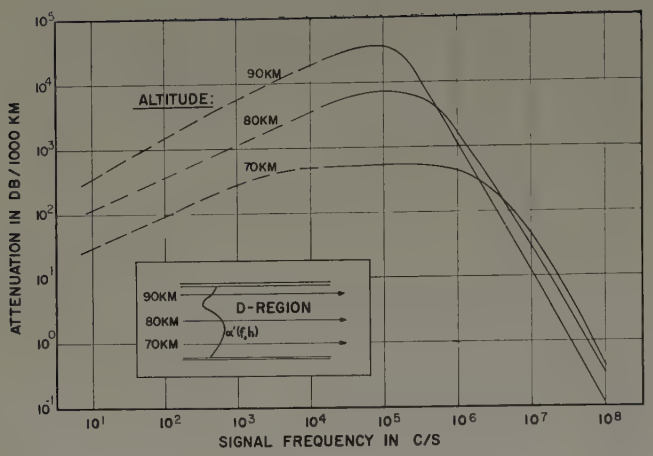


FIG. 8. Another example of large cumulative effects due to the weakly conducting D-Region. For long horizontal paths through the D-Region losses can run greater than 10,000 *db* at frequencies of 100 KC. Low frequency radio astronomy and radio probing of the interplanetary medium will be made possible by space stations away from the earth's obscuring ionospheric blanket.

Summary

Some of the basic electrical properties of the D-region have been evaluated with a major emphasis on their nature over the wide frequency range of 10 to 10^8 c/s. A number of prominent effects appear which help to fill in the picture of significant features of the electromagnetic environment of spacecraft. To sum up broadly, these properties and effects are as follows:

D-Region Properties

The region occupies the lower 25 *km* of the ionosphere. Electron densities are relatively low and vary only some twentyfold. However, the energy dissipating factor, the electron collision frequency, is quite large and has a one hundredfold range in magnitude. This latter characteristic is generally the most dominant factor in D-region effects.

Table 3 gives an orders-of-magnitude summary of basic properties of the D-region. The electrical conductivity is relatively low being only some 10^{-4} times that of sea water at low frequencies. However, it varies through a range of some 10^{10} depending on frequency. The dielectric "constant" is fairly constant but develops negative values at the top of the D-region. Although low in value the conductivity of the region is the dominant factor in refractive properties at very low frequencies, making the region quite shiny then.

D-Region Effects

Perhaps the most important effect of the D-region on spacecraft electrical environment is in regard to radio communications. For long paths through the region certain unfavorable orientations of signal wavefronts with the geomagnetic field can lead to extremely high signal loss. At frequencies of 100 mc and greater, effects should generally disappear. However, even at 100 mc wave amplitudes may occasionally be reduced ten and possibly a hundredfold on a geophysicall disturbed day.

TABLE 3

Range in Electrical Properties of D-Region

Altitude, H, in km..... 64–91
Electron density, N, in e/m^3..... $10^8 - 3 \times 10^9$
Electron collision frequency, ν, in sec^{-1}. $4 \times 10^7 - 4 \times 10^5$

Electrical Property	Decade range (min–max) at frequency of:					
	10^2 c/s		10^5 c/s		10^8 c/s	
	min	max	min	max	min	max
Conductivity, σ, in (ohm-met)$^{-1}$	10^{-7}	10^{-4}	10^{-7}	10^{-4}	10^{-11}	10^{-}
Dielectric constant, ϵ_r	-1	1	-0.1	1	1	1
Index of refraction, n	1	100	1	10	1	1
Local attenuation factor, α', in db/km	10^{-1}	10^{0}	10^{-1}	10^{1}	10^{-4}	10^{-}
Horizontal attenuation, α'_h, in db/1000 km	10^2	10^3	10^2	10^4	10^{-1}	10^{0}
Vertical attenuation, α'_v, in db	10^1		10^2		10^{-2}	

References

[1] J. W. BOND, Plasma Physics and Hypersonic Flight, *J. Propulsion*, Vol. 28, pp. 229–235, 1958.

[2] P. A. GOLDBERG, Electrical Properties of Hypersoni Shock Waves and Their Effect on Aircraft Radio an Radar, *Boeing Airplane Company Document D2-199 Armed Serv. Tech. Info. Agency Doc. AD 135-058*, 5 p., 1957.

[3] J. R. WAIT, An Extension to the Mode Theory of VL Propagation, *J. Geophys. Res.*, Vol. 63, pp. 125–13 1958.

[4] L. R. O. STOREY, An Investigation of Whistling A mospherics, *Phil. Trans. Roy. Soc., London, Ser. A* Vol. 246, pp. 113–141, 1953.

[5] R. A. HELLIWELL, J. H. CRARY, J. H. POPE AND R. I SMITH, The Nose Whistler, A New High Latitud Phenomenon, *J. Geophys. Res.*, Vol. 61, pp. 139–142 1956.

[6] P. A. GOLDBERG, Electromagnetic Phenomena of Natur Origin in the 1.0 to 150 c/s Band, *Nature*, Vol. 177 pp. 1219–1221, 1956.

[7] P. A. GOLDBERG, Worldwide Lightning Fields and Mag netic Pulsations in the 1 to 150 c/s Band, *IRE Trans Ant. and Propgn.*, Vol. 5, p. 165, 1957.

[8] H. POEVERLEIN, Low-Frequency Reflection in th Ionosphere, *J. Atmos. and Terr. Phys.*, Vol. 12, pp 126–139, 1958.

[9] R. J. NERTNEY, The Lower E and D Region of th Ionosphere as Deduced from Long Radio Wave Meas urements, *J. Atmos. and Terr. Phys.*, Vol. 3, pp. 92–107 1953.

[10] M. NICOLET, The Collision Frequency of Electrons in th Ionosphere, *J. Atmos. and Terr. Phys.*, Vol. 3, pp. 201 211, 1953.

[11] J. A. STRATTON, *Electromagnetic Theory*, Chap. V, Mc Graw-Hill Book Co., New York, 1941.

[12] S. K. MITRA, *The Upper Atmosphere*, Chap. VI, Th Asiatic Society, Calcutta, 1952.

[13] J. C. SEDDON, Performance of a Rocket Borne 7.5 M Transmitting Antenna in the Ionosphere, *Nav. Res. Lab. Upper Atmos. Rpt. No. 28*, 1957.

[14] S. K. MITRA, op. cit. p. 628.

Format of Technical Papers for AAS Journal

The editors will appreciate the cooperation of authors in using the following directions for the preparation of manuscripts. These directions have been compiled with a view toward eliminating unnecessary correspondence, avoiding the return of papers for changes, and reducing the charges made for "author's corrections."

Manuscripts

Papers should be submitted in original typewriting (if possible) on one side only of white paper sheets, and should be double or triple spaced with wide margins. However, good quality reproduced copies (e.g. multilith) are acceptable. An additional copy of the paper will facilitate review.

Company Reports

The paper should not be merely a company report. If such a report is to be used as the basis for the paper, appropriate changes should be made in the title page. Lists of figures, tables of contents, and distribution lists should all be deleted.

Titles

The title should be brief, but express adequately the subject of the paper. A footnote reference to the title should indicate any meeting at which the paper has been presented. The name and initials of the author should be written as he prefers; all titles and degrees or honors will be omitted. The name of the organization with which the author is associated should be given in a separate line to follow his name.

Abstracts

An abstract should be provided, preceding the introduction, covering contents of the paper. It should not exceed 200 words.

Headings

The paper can be divided into principal sections as appropriate. Headings or paragraphs are not numbered.

Mathematical Work

As far as possible, formulas should be typewritten. Greek letters and other symbols not available on the typewriter should be carefully inserted in ink. Each such symbol should be identified unambiguously the first time it appears. The distinction between capital and lower-case letters should be clearly shown. Avoid confusion between zero (0) and the letter *O*; between the numeral (1), the letter l, and the prime (′); between alpha and *a*, kappa and *k*, mu and *u*, nu and *v*, eta and *n*.

The level of subscripts, exponents, subscripts to subscripts, and exponents in exponents should be clearly indicated.

Complicated exponents and subscripts should be avoided when possible to represent by a special symbol.

Fractions in the body of the text and fractions occurring in the numerators or denominators of fractions should be written with the solidus. Thus

$$\frac{\cos(\pi x/2b)}{\cos(\pi\alpha/2b)}$$

is the preferred usage.

The intended grouping of handwritten formulas can be made clear by slight variations in spacing, but this procedure is not acceptable in printed formulas. To avoid misunderstanding, the order of symbols should therefore be carefully considered. Thus

$(a + bx)\cos t$ is preferable to $\cos t\,(a + bx)$

In handwritten formulas the size of parentheses, brackets and braces can vary more widely than in print. Particular attention should therefore be paid to the proper use of parentheses, brackets, and braces (which should be used in this order). Thus

$$\{[a + (b + cx)^n]\cos ky\}^2$$

is required rather than $((a + (b + cx)^n)\cos ky)^2$.

Equations are numbered and referred to in text as (15).

Illustrations

Drawings should be made with black India ink on white paper or tracing cloth, and should be at least double the desired size of the cut. Each figure number should be marked with soft pencil in the margin or on the back of the drawing. The width of the lines of such drawings and the size of the lettering must allow for the necessary reduction. Reproducible glossy photographs are acceptable. However, drawings which are unsuitable for reproduction will be returned to the author for redrawing. Legends accompanying the drawings should be typewritten on a separate sheet, properly identified.

Bibliography

References should be grouped together in a bibliography at the end of the manuscript. References to the bibliography should be made by numerals between square brackets [4].

The following examples show the approved arrangements:

for books—[1] HUNSAKER, J. C. and RIGHTMIRE, B. S., *Engineering Applications of Fluid Mechanics*, McGraw-Hill Book Co., New York, 1st ed., 1947, p. 397.

for periodicals—[2] Singer, S. F., "Artificial Modification of the Earth's Radiation Belt," *J. Astronaut. Sci.*, *6* (1959), 1–10.